

Performance of Self-Report to Establish Cancer Diagnoses in Disaster Responders and Survivors, World Trade Center Health Registry, New York, 2001–2007

JIEHUI LI, MBBS, MS^a
JAMES E. CONE, MD, MPH^a
ABIGAIL K. ALT, MD^b
DAVID R. WU, MS^a
JONATHAN M. LIFF, PhD^c
MARK R. FARFEL, ScD^a
STEVEN D. STELLMAN, PhD,
MPH^{a,d}

ABSTRACT

Objective. Large-scale disasters may disrupt health surveillance systems, depriving health officials and researchers of timely and accurate information needed to assess disaster-related health effects and leading to use of less reliable self-reports of health outcomes. In particular, ascertainment of cancer in a population is ordinarily obtained through linkage of self-reported data with regional cancer registries, but exclusive reliance on these sources following a disaster may result in lengthy delays or loss of critical data. To assess the impact of such reliance, we validated self-reported cancer in a cohort of 59,340 responders and survivors of the World Trade Center disaster against data from 11 state cancer registries (SCRs).

Methods. We focused on residents of the 11 states with SCRs and on cancers diagnosed from September 11, 2001, to the date of their last survey participation. Medical records were also sought in a subset of 595 self-reported cancer patients who were not recorded in an SCR.

Results. Overall sensitivity and specificity of self-reported cancer were 83.9% (95% confidence interval [CI] 81.9, 85.9) and 98.5% (95% CI 98.4, 98.6), respectively. Site-specific sensitivities were highest for pancreatic (90.9%) and testicular (82.4%) cancers and multiple myeloma (84.6%). Compared with enrollees with true-positive reports, enrollees with false-negative reports were more likely to be non-Hispanic black (adjusted odds ratio [aOR] = 1.8, 95% CI 1.2, 2.9) or Asian (aOR=2.2, 95% CI 1.2, 4.1). Among the 595 cases not recorded in an SCR, 13 of 62 (21%) cases confirmed through medical records were reportable to SCRs.

Conclusion. Self-report of cancer had relatively high sensitivity among adults exposed to the World Trade Center disaster, suggesting that self-reports of other disaster-related conditions less amenable to external validation may also be reasonably valid.

^aNew York City Department of Health and Mental Hygiene, World Trade Center Health Registry, Long Island City, NY

^bSwedish Medical Center, Seattle, WA

^cEmory University, Rollins School of Public Health, Atlanta, GA

^dColumbia University, Mailman School of Public Health, Department of Epidemiology, New York, NY

Address correspondence to: Jiehui Li, MBBS, MS, New York City Department of Health and Mental Hygiene, World Trade Center Health Registry, 42-09 28th St., 7th Fl., Long Island City, NY 11101; tel. 347-396-2868; fax 347-396-2893; e-mail <jli3@health.nyc.gov>.

©2016 Association of Schools and Programs of Public Health

Identification and tracking of long-term physical and mental health effects of man-made and natural disasters often rely on self-reports of clinically diagnosed health conditions obtained longitudinally through questionnaires,¹ but the accuracy of such self-reported diagnoses is often difficult to assess because of large sample sizes and limited funding. Accuracy of self-reported health data can depend on such factors as type of condition, study population demographics, and recall period length.²⁻⁹

Linkage of cohort data with independent cancer registries enables assessment of cancer self-report accuracy and may reflect self-report accuracy of other endpoints for which no outside data are available. The sensitivity of self-reported cancer is reported to vary by site, treatment, and number of previous tumors.^{6,10-12} False-negative self-report of cancer diagnoses determined by comparison with cancer registry data is associated with older age,^{6,11,13} nonwhite race, increased time since cancer diagnosis,⁶ lower education level,^{2,8} male sex, and urban living.²

The World Trade Center (WTC) Health Registry (WTCHR) has followed a cohort of 71,434 people since 2003 to identify and track the long-term health effects of the September 11, 2001 (hereinafter 9/11), terrorist attacks through surveys.¹⁴ Although the WTCHR periodically requests to compare its data with those of other health registries, such as state cancer registries (SCRs) and the Statewide Planning and Research Cooperative System, the time lags between diagnosis and data availability for comparison often limits timely surveillance of emerging or rare health conditions. Thus, the accuracy of self-reported health information, including cancer diagnosis, provided by WTCHR enrollees at enrollment and in subsequent waves is important. We examined the performance of self-reported health data collected by the WTCHR, using cancer as an example, by comparing WTCHR survey data with data obtained via linkage with SCRs to validate self-reported cancer diagnosis. We also examined correlates of false-negative and false-positive reports.

METHODS

The World Trade Center Health Registry

The WTCHR is a cohort study of 71,434 people who were directly exposed to the destruction of the WTC and surrounding buildings and its aftermath. Details of eligibility and recruitment are available elsewhere.^{14,15} In brief, people with potential exposure were recruited from lists of businesses and building occupants and via public outreach and media campaigns. Potential enrollees were screened for eligibility, with those enrolled

belonging to one or more of the following groups: rescue/recovery workers and volunteers, lower Manhattan residents, area workers, passersby, schoolchildren, and school staff members. Baseline data (Wave 1) were gathered in 2003–2004 via computer-assisted telephone interview (95%) or in-person interview (5%). Wave 1 data included data on demographics, current health status, medical history, and exposure. An adult follow-up survey (Wave 2) conducted from November 2006 through December 2007 (response rate: 68%, $n=46,602/68,959$) obtained additional exposure data and updated health information.¹⁶ The final cohort ($n=71,434$) consisted of 30,664 (42.9%) responders (rescue/recovery workers and volunteers) and 40,770 (57.1%) survivors (residents, passersby, area workers, schoolchildren, and school staff members in Lower Manhattan on the morning of 9/11).

Study sample

The study sample for this analysis was limited to adult enrollees aged ≥ 18 years at enrollment (Wave 1) who completed cancer questions in Wave 1 or follow-up (Wave 2) surveys. The study was also limited to residents since 9/11 of the 11 states (California, Connecticut, Florida, Massachusetts, New Jersey, New York, North Carolina, Ohio, Pennsylvania, Texas, and Washington) in which we conducted cancer record linkage. We excluded proxies and withdrawals and included only cancers—reported either by enrollees or the SCRs—that were diagnosed between September 12, 2001, and the date of their last survey participation (Wave 1 or Wave 2). A total of 59,340 enrollees met inclusion criteria (Figure 1).

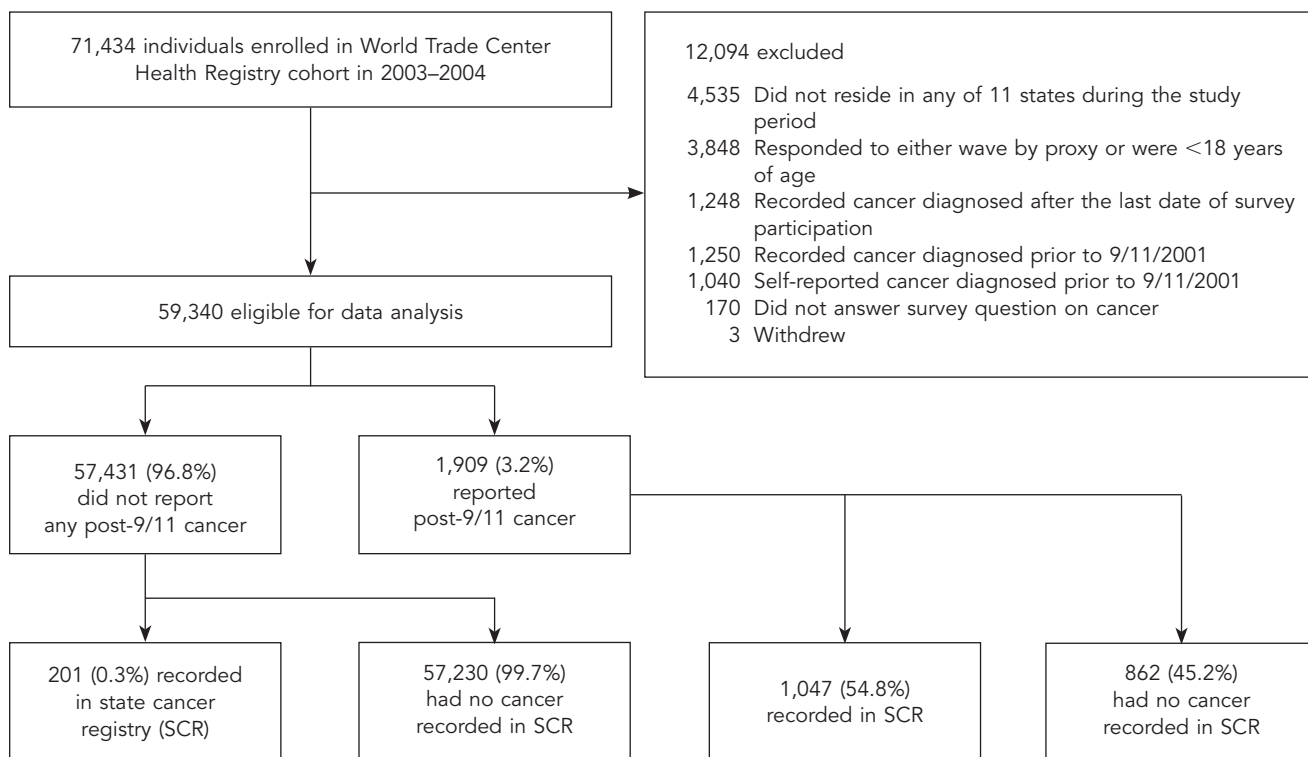
Self-reported cancer

In Wave 1, we asked (1) “Have you ever been told by a doctor or other health professional that you had cancer or a malignancy of any kind?”, (2) “Did a doctor or other health professional first tell you that you had cancer or a malignancy of any kind before 9/11 or after 9/11?”, and (3) “What kind of cancer was it?” with a drop-down menu of cancer sites. Wave 2 had similar questions, but the question about cancer site required an open-ended response and inquired about year of diagnosis. A self-reported post-9/11 cancer was defined as a positive answer to the first cancer question and a reported diagnosis after 9/11.

Cancer identified from cancer registry linkages

We matched enrollees with the people registered in the 11 SCRs, all of which adopted Link Plus, a probabilistic record linkage program developed by the Centers for Disease Control and Prevention.¹⁷ We provided full

Figure 1. Flow diagram of people enrolled in the World Trade Center Health Registry cohort who self-reported cancer compared with records in a state cancer registry, 2003–2004^a



^aRefers to people who resided in any of the 11 states for which matched state cancer registry data were obtained: California, Connecticut, Florida, Massachusetts, New Jersey, New York, North Carolina, Ohio, Pennsylvania, Texas, and Washington.

SCR = state cancer registry

name, sex, race/ethnicity, birth date, complete address of residence, and social security number when available to each SCR. Matches were reported by each SCR; also reported by each SCR was information on primary cancer site(s), histology, stage, diagnosis date, and state where cancer was diagnosed. Linked cancer data from all SCRs were available through December 31, 2008. Cancer site was defined according to the International Classification of Diseases for Oncology, Third Edition, and grouped by using the Surveillance Epidemiology and End Results (SEER) site recode codes for primary site and histology.¹⁸

Medical record confirmation of false-positive report

We investigated self-reports of cancer that were not also confirmed by the SCRs by contacting enrollees' physicians for confirmation of cancer diagnoses. This investigation required first contacting the enrollees for permission to communicate with their physicians and

then reviewing relevant medical reports and records. Because of limited resources, we carried out this investigation from June 2009 to January 2010 only among the 595 enrollees then living in New York State.

Data analysis

We evaluated the performance of self-report using sensitivity, specificity, and positive predictive value (PPV). We defined a true-positive report of cancer as a self-report of cancer recorded in an SCR, a false-negative report of cancer as non-self-report of cancer that was recorded in an SCR, a false-positive report of cancer as a self-report not recorded in an SCR, and a true-negative report of cancer as a non-report and non-recorded cancer (Figure 2). We defined sensitivity as the proportion of true-positive reports among recorded cancers. We defined specificity as the proportion of true-negative reports among non-recorded cancers. We defined positive predictive value (PPV) (i.e., agreement

of self-report with SCR) as the proportion of true-positive reports among all self-reports (true-positive reports and false-positive reports).

A matched cancer site refers to verification of the anatomical site of the self-reported cancer using SEER site recode rules. The term false-positive is used for consistency with standard validation analyses and does not imply that those reporting cancer actually do not have cancer. We recognize that a cancer may not have been reported to the SCR for various reasons. Understanding that a matched individual may not match to the cancer site, we assessed the performance of self-report separately, by individual and by site.

We performed bivariate and multivariable analyses using logistic regression modeling to assess whether enrollees' sociodemographic characteristics, medical history, or number of cancer sites were associated with either false-negative or false-positive reports. We computed unadjusted odds ratios (ORs), adjusted odds ratios (aORs), and 95% confidence intervals (CIs). The OR represents the odds of having a false-negative or false-positive report given the presence of a variable of interest, compared with the odds of having a false-negative or false-positive report given the absence of a particular variable of interest. We computed aORs and 95% CIs using a multivariate model in which variables such as sociodemographics, Wave 2 participation, history of other medical conditions, and probable posttraumatic stress disorder (PTSD) were included if they were significantly associated with false-negative or false-positive reporting in bivariate analyses. We evaluated socioeconomic status as a combination of annual household income and education at three levels: low (education ≤ 12 years and annual household income $< \$25,000$), high ($\geq$ college degree and annual household income $\geq \$50,000$), and intermediate (between low and high socioeconomic status). In analyzing people with false-positive reports, we excluded enrollees who reported only non-melanoma or unspecified

skin cancers because these tumors are not reportable to SCRs.

We also examined whether or not a positive response to the Wave 1 cancer question was consistent with the Wave 2 response by computing percentages and 95% CIs of agreement. We performed all data analyses using SAS[®] version 9.2.¹⁹

RESULTS

Sensitivity and specificity of self-report

Overall sensitivity and specificity of self-report were 83.9% (95% CI 81.9, 85.9) and 98.5% (95% CI 98.4, 98.6), respectively. Sensitivity was greater among participants in Wave 1 and Wave 2 than among participants in Wave 1 only (87.5% vs. 67.3%); however, specificity was not substantially different between these two groups (Table 1).

Of 1,909 enrollees who reported a cancer diagnosis, 1,393 (73.0%) specified a single cancer site, 80 (4.2%) reported two or more sites, and 436 (22.8%) did not specify any site. Sensitivity varied by site. The highest site-specific sensitivities for self-report of cancer were observed for pancreatic (90.9%), multiple myeloma (84.6%), and testicular (82.4%) cancers. The highest PPV of self-reported cancer was for multiple myeloma (100.0%), followed by prostate (93.5%) and testicular (93.3%) cancers, while the lowest was for melanoma of the skin (40.5%) (Table 1).

Factors associated with false-negative report

Of the 1,248 enrollees with a primary cancer site recorded in an SCR, 201 (16.1%) did not report having any cancer diagnosed in either wave (Table 1). Compared with the 1,047 enrollees with true-positive reports, enrollees with false-negative reports were more likely to be Wave 2 nonparticipants (aOR=2.9, 95% CI 2.0, 4.1), non-Hispanic black (aOR=1.8, 95% CI 1.2, 2.9) or Asian (aOR=2.2, 95% CI 1.2, 4.1), or have not

Figure 2. Definitions of measures used to assess the performance of self-report of cancer among people enrolled in the World Trade Center Health Registry^a compared with records in a state cancer registry, 2001–2007^b

Self-report status of cancer diagnosis	Record status in a state cancer registry		
	Recorded		Not recorded
	Self-reported	True-positive report (TPR)	False-positive report (FPR)
	Not self-reported	False-negative report (FNR)	True-negative report (TNR)

^aRefers to people who resided in any of the 11 states for which matched state cancer registry data were obtained: California, Connecticut, Florida, Massachusetts, New Jersey, New York, North Carolina, Ohio, Pennsylvania, Texas, and Washington.

^bSensitivity of self-report = $TPR / (TPR + FNR)$; specificity of self-report = $TNR / (FPR + TNR)$; positive predictive value of self-report = $TPR / (TPR + FPR)$

Table 1. Sensitivity, specificity, and positive predictive values of self-reported cancer compared with records in a state cancer registry, World Trade Center Health Registry, New York, 2001–2007

Characteristic	Number of cancers recorded in state cancer registry			Number of cancers not recorded in state cancer registry			Percent sensitivity of self-report (95% CI)	Percent specificity of self-report (95% CI) ^a	Percent PPV of self-report (95% CI)
	Total	Self-reported (TPR)	Not self-reported (FNR)	Total	Self-reported (FPR)	Not self-reported (TNR)			
By individual									
All	59,340	1,248	201	58,092	862 ^b	57,230	83.9 (81.9, 85.9)	98.5 (98.4, 98.6)	54.9 (52.6, 57.1)
Enrollees participating in Wave 1 ^c only	19,809	223	73	19,586	83	19,503	67.3 (61.1, 73.4)	99.6 (99.5, 99.7)	64.4 (58.2, 70.5)
Enrollees participating in Wave 1 and Wave 2 ^c	39,531	1,025	128	38,506	779	37,727	87.5 (85.5, 89.5)	98.0 (97.8, 98.1)	53.5 (51.1, 55.9)
By selected cancer site ^d									
Brain and other nervous system	59,340	44	30	59,296	10	59,286	31.8 (18.1, 45.6)	99.9	58.3 (38.6, 78.1)
Colon and rectum	59,340	77	31	59,263	28	59,235	59.7 (48.8, 70.7)	99.9	62.2 (51.1, 73.2)
Corpus uterus and not otherwise specified	59,340	33	12	59,307	11	59,296	63.6 (47.2, 80.1)	99.9	65.6 (49.2, 82.1)
Female breast	23,255	233	58	23,022	28	22,994	75.1 (69.6, 80.7)	99.9	86.2 (81.5, 91.0)
Kidney and renal pelvis	59,340	46	19	59,294	8	59,286	58.7 (44.5, 72.9)	99.9	77.1 (63.2, 91.1)
Lung and bronchus	59,340	78	26	59,262	16	59,246	66.7 (56.2, 77.1)	99.9	76.5 (66.4, 86.6)
Lymphoma, leukemia, or blood cancer	59,340	122	28	59,218	37	59,181	77.1 (69.6, 84.5)	99.9	71.8 (64.1, 79.5)
Melanoma of the skin	59,340	99	48	59,241	75	59,166	51.5 (41.7, 61.4)	99.9	40.5 (31.9, 49.1)
Multiple myeloma	59,340	13	2	59,327	0	59,327	84.6 (65.0, 100.0)	100.0	100.0 (100.0, 100.0)
Oral cavity and pharynx	59,340	27	21	59,313	5	59,308	22.2 (6.5, 37.9)	99.9	54.6 (25.1, 84.0)
Ovary	59,340	16	8	59,324	11	59,313	50.0 (25.5, 74.5)	99.9	42.1 (19.9, 64.3)
Pancreas	59,340	11	1	59,329	2	59,327	90.9 (73.9, 100.0)	100.0	83.3 (62.3, 100.0)
Prostate	36,085	247	59	35,838	13	35,825	76.1 (70.8, 81.4)	99.9	93.5 (90.1, 96.9)
Stomach	59,340	13	7	59,327	6	59,321	46.2 (19.1, 73.3)	99.9	50.0 (21.7, 78.3)
Testis	59,340	17	3	59,323	1	59,322	82.4 (64.2, 100.0)	100.0	93.3 (80.7, 100.0)
Thyroid	59,340	60	16	59,280	19	59,261	73.3 (62.1, 84.5)	99.9	69.8 (58.5, 81.2)
Urinary bladder	59,340	47	17	59,293	7	59,286	63.8 (50.1, 77.6)	99.9	81.1 (68.5, 93.7)

^aThe 95% lower confidence limits for percent specificity of self-report by cancer site were above 99%; therefore the 95% CIs for specificity of self-report by cancer site are not shown.

^bOf 862 enrollees with false-positive reports, 390 reported non-melanoma or unspecified skin cancer only.

^cWave 1 referred to baseline survey at enrollment in 2003–2004, and Wave 2 referred to the first follow-up survey in 2006–2007.

^dThe sum of numbers from each cancer site listed in the table is not equal to the total number of eligible enrollees in the corresponding column because only cancer sites with ≥10 reports are presented in the table, and enrollees may have reported more than one cancer site.

TPR = true-positive report

FNR = false-negative report

FPR = false-positive report

TNR = true-negative report

PPV = positive predictive value

CI = confidence interval

provided a social security number (aOR=1.6, 95% CI 1.0, 2.6) (Table 2). Enrollee characteristics such as age, sex, smoking status, history of cardiovascular disease, emphysema, diabetes, or PTSD were not associated with false-negative reports. A non-English-language Wave 1 interview was strongly associated with false-negative report in bivariate analyses (OR=3.7, 95% CI 2.0, 6.9). However, this variable was excluded from multivariable analyses because of strong collinearity with socioeconomic status.

Factors associated with false-positive report

A total of 1,909 enrollees reported having post-9/11 cancer diagnosed during the study period, of whom 862 (45.2%) were not recorded in an SCR. Of these 862 enrollees, 31 (3.6%) reported more than one cancer and 390 (45.2%) reported non-melanoma or unspecified skin cancer that was not reportable to an SCR.

After excluding the 390 non-reportable non-melanomas or unspecified skin cancer cases, we compared 472 enrollees who had false-positive reports with 1,047 enrollees who had true-positive reports (Table 3). Several factors that were not associated with false-negative reports were associated with false-positive reports as compared with true-positive reports. Enrollees with false-positive reports were more likely than enrollees with true-positive reports to be younger (aOR=3.6, 95% CI 1.0, 13.6 for adults aged 18–24 years; aOR=2.3, 95% CI 1.8, 3.0 for adults aged 25–44 years), current smokers than never smokers (aOR=1.9, 95% CI 1.3, 2.6), and to have probable PTSD than not (aOR=1.7, 95% CI 1.3, 2.3) in the adjusted model. Factors such as race/ethnicity and providing a social security number that were associated with false-negative reports were not associated with false-positive reports. Those who did not participate in Wave 2 were less likely than those who did to have false-positive reports (aOR=0.6, 95% CI 0.4, 0.8).

Medical record confirmation

In the sub-analysis of 595 New York State residents for whom we sought medical record confirmation of self-reported cancer, 260 enrollees (43.7%) responded to our investigation; 138 (53.1%) claimed they had never had a cancer diagnosis, 92 (35.4%) gave permission for us to obtain further information from their physician, and 30 (11.5%) refused to give permission. Among those who gave permission, 62 self-reported cancer cases (67.4%) were physician confirmed, 19 (20.7%) were physician confirmed as not having cancer, eight (8.7%) were not confirmed because the treating physicians failed to provide information, and three (3.3%) were not confirmed because medical records

could not be located. Among the 62 self-reported cancers confirmed by primary or treating physicians, 13 were reportable cancer cases and the remaining were non-reportable cancers because they were either non-melanoma skin cancer or benign tumors.

Consistency of self-report of cancer in both waves

When asked at Wave 1, “Have you ever been told by a doctor or other health professional that you had cancer or a malignancy of any kind?”, 611 of 39,531 people who participated in both Wave 1 and Wave 2 answered affirmatively. When asked the same question in Wave 2, 560 (91.7%) of these 611 enrollees again answered affirmatively, and 51 (8.3%) provided inconsistent answers. The cancer linkage proportion was 60.9% ($n=341/560$) for enrollees who provided consistent answers and 17.6% ($n=9/51$) for those who provided inconsistent answers.

DISCUSSION

This study examined the performance of self-reported cancer diagnoses in a 9/11-exposed cohort and identified correlates of false-negative and false-positive reporting. Sensitivity of self-reported cancer among those who responded to both surveys was relatively high (87.5%). It was also higher than in several other U.S. studies using cancer registry data as the gold standard (60.8%–74.2%)^{3,6,20} and similar studies conducted in other countries (40.0%–57.5%).^{8,13,21} The relatively high sensitivity in this study may be attributable to the use of multiple SCRs, to giving participants an opportunity to report a diagnosis twice (during Wave 1 and Wave 2), or to the relatively short recall period (2–6 years) compared with other studies (>20 years).^{6,11}

The site-specific sensitivity varied considerably and was low for cancers in the oral cavity and the pharynx (22.2%) or brain and other nervous system (31.8%), indicating substantial underreporting of these cancer sites in the study sample. False-negative reporting among non-Hispanic black and Asian patients and those who did not provide a social security number was also prevalent. Similar findings were reported in a community-based study where the most often underreported cancer sites were central nervous system and lip, oral cavity, and pharynx, and nonwhite patients were 10 times more likely than white patients to provide a false-negative report of their cancer history.⁶

Knowing the patterns of false-negative reporting is important for surveillance of cancer incidence among WTC/HR enrollees because high validity of self-reported cancer data collected from follow-up surveys may partially compensate for the delay in time from

Table 2. Predictors of false-negative self-report and true-positive self-report of cancer diagnosis compared with records in a state cancer registry, World Trade Center Health Registry, New York, 2001–2007

Characteristic	False-negative self-report of cancer Number (percent) ^a	True-positive self-report of cancer Number (percent) ^a	Unadjusted odds ratio ^b (95% CI)	Adjusted ^{b,c} odds ratio (95% CI)
Total	201 (100.0)	1,047 (100.0)		
Participated in Wave 2 ^d				
No	73 (36.3)	150 (14.3)	3.4 (2.4, 4.8)	2.9 (2.0, 4.1)
Yes	128 (63.7)	897 (85.7)	Ref.	Ref.
Age at enrollment, in years				
18–24	1 (0.5)	4 (0.4)	1.3 (0.1, 11.4)	1.1 (0.1, 11.4)
25–44	38 (18.9)	257 (24.5)	0.7 (0.5, 1.1)	0.8 (0.5, 1.1)
45–64	131 (65.2)	663 (63.3)	Ref.	Ref.
≥65	30 (14.9)	123 (11.7)	1.2 (0.8, 1.9)	1.0 (0.6, 1.7)
Unknown ^e	1 (0.5)	0 (0.0)	NA	NA
Sex				
Male	122 (60.7)	595 (56.8)	1.2 (0.9, 1.6)	NA
Female	79 (39.3)	452 (43.2)	Ref.	Ref.
Race/ethnicity				
Non-Hispanic white	116 (57.7)	783 (74.8)	Ref.	Ref.
Non-Hispanic black	34 (16.9)	113 (10.8)	2.0 (1.3, 3.1)	1.8 (1.2, 2.9)
Hispanic	20 (10.0)	72 (6.9)	1.9 (1.1, 3.2)	1.6 (0.9, 2.8)
Asian	21 (10.4)	44 (4.2)	3.2 (1.8, 5.6)	2.2 (1.2, 4.1)
American Indian/Alaska Native	3 (1.5)	6 (0.6)	3.4 (0.8, 13.7)	3.5 (0.8, 14.7)
Other or unknown	7 (3.5)	29 (2.8)	1.6 (0.7, 3.8)	1.1 (0.5, 2.8)
Socioeconomic status at enrollment				
Low (<grade 12 and <\$25,000 annual household income)	13 (6.5)	16 (1.5)	5.2 (2.4, 11.4)	2.1 (0.8, 5.1)
Intermediate	127 (63.2)	640 (61.1)	1.3 (0.9, 1.8)	1.1 (0.7, 1.5)
High (≥college degree and ≥\$50,000 annual household income)	61 (30.3)	391 (37.3)	Ref.	Ref.
Provided social security number				
Yes (full or partial)	168 (83.6)	960 (91.7)	Ref.	Ref.
No	33 (16.4)	87 (8.3)	2.2 (1.4, 3.3)	1.6 (1.0, 2.6)
Language in which the enrollment interview was administered				
English	183 (91.0)	1,020 (97.4)	Ref.	Ref.
Other	18 (9.0)	27 (2.6)	3.7 (2.0, 6.9)	NA
Smoking status at enrollment				
Current	24 (11.9)	106 (10.1)	1.1 (0.7, 1.9)	NA
Former	74 (36.8)	411 (39.3)	0.9 (0.7, 1.2)	NA
Never	103 (51.2)	516 (49.3)	Ref.	Ref.
Unknown	0 (0.0)	14 (1.3)	NA	NA
Any history of medical condition, excluding cancer, reported at enrollment ^f				
Yes	70 (34.8)	363 (34.7)	1.0 (0.7, 1.4)	NA
No	131 (65.2)	684 (65.3)	Ref.	Ref.
Probable PTSD (PTSD checklist score ≥44) at enrollment				
Yes	36 (17.9)	174 (16.6)	1.1 (0.7, 1.6)	NA
No	163 (81.1)	860 (82.1)	Ref.	Ref.
Unknown	2 (1.0)	13 (1.2)	0.8 (0.2, 3.6)	NA
Number of cancer sites identified by cancer registry				
1	191 (95.0)	962 (91.9)	1.7 (0.9, 3.3)	NA
≥2	10 (5.0)	85 (8.1)	Ref.	Ref.

^aPercentages may not total to 100 because of rounding.

^bThe OR represents the odds of having a false-negative or false-positive report given the presence of a variable of interest, compared with the odds of having a false-negative or false-positive report given the absence of a particular variable of interest.

^cAdjusting for variables in this table that were significant or marginally significant in bivariate analyses

^dWave 2 refers to the first follow-up survey in 2006–2007.

^eUnknown age but indicated at ≥18 years at enrollment

^fHistory of cardiovascular disease (heart disease, hypertension, angina, or heart attack), emphysema, stroke, or diabetes

CI = confidence interval

Ref. = reference group

NA = not applicable

PTSD = posttraumatic stress disorder

Table 3. Predictors of false-positive self-report and true-positive self-report of cancer compared with records in a state cancer registry, World Trade Center Health Registry, New York, 2001–2007^a

Characteristic	False-positive self-report of cancer ^a Number (percent) ^b	True-positive self-report of cancer Number (percent) ^b	Unadjusted OR (95% CI)	Adjusted ^c OR (95% CI)
Total	472 (100.0)	1,047 (100.0)		
Participated in Wave 2 ^d				
No	47 (10.0)	150 (14.3)	0.7 (0.5, 0.9)	0.6 (0.4, 0.8)
Yes	425 (90.0)	897 (85.7)	Ref.	Ref.
Age at enrollment, in years				
18–24	6 (1.3)	4 (0.4)	4.4 (1.2, 15.6)	3.6 (1.0, 13.6)
25–44	199 (42.2)	257 (24.5)	2.3 (1.8, 2.9)	2.3 (1.8, 3.0)
45–64	228 (48.3)	663 (63.3)	Ref.	Ref.
≥65	38 (8.1)	123 (11.7)	0.9 (0.6, 1.3)	0.9 (0.6, 1.4)
Unknown ^e	1 (0.2)	0 (0.0)	NA	NA
Sex				
Male	265 (56.1)	595 (56.8)	1.0 (0.8, 1.2)	NA
Female	207 (43.9)	452 (43.2)	Ref.	Ref.
Race/ethnicity				
Non-Hispanic white	342 (72.5)	783 (74.8)	Ref.	Ref.
Non-Hispanic black	44 (9.3)	113 (10.8)	0.9 (0.6, 1.3)	0.8 (0.6, 1.2)
Hispanic	49 (10.4)	72 (6.9)	1.6 (1.1, 2.3)	1.2 (0.8, 1.8)
Asian	24 (5.1)	44 (4.2)	1.2 (0.7, 2.1)	1.1 (0.6, 1.9)
American Indian/Alaska Native	5 (1.1)	6 (0.6)	1.9 (0.6, 6.3)	1.7 (0.5, 5.8)
Unknown	8 (1.7)	29 (2.8)	0.6 (0.3, 1.4)	0.7 (0.3, 1.7)
Socioeconomic status at enrollment				
Low (<grade 12 and <\$25,000 annual household income)	13 (2.8)	16 (1.5)	2.1 (1.0, 4.4)	2.2 (1.0, 5.1)
Intermediate	306 (64.8)	640 (61.1)	1.2 (1.0, 1.5)	1.1 (0.9, 1.5)
High (≥college degree and ≥\$50,000 annual household income)	153 (32.4)	391 (37.3)	Ref.	Ref.
Provided social security number				
Yes (full or partial)	434 (91.9)	960 (91.7)	Ref.	Ref.
No	38 (8.1)	87 (8.3)	1.0 (0.6, 1.4)	NA
Smoking status at enrollment				
Current smoker	95 (20.1)	106 (10.1)	2.0 (1.5, 2.8)	1.9 (1.3, 2.6)
Former smoker	145 (30.7)	411 (39.3)	0.8 (0.6, 1.0)	1.0 (0.7, 1.3)
Never smoker	229 (48.5)	516 (49.3)	Ref.	Ref.
Unknown	3 (0.6)	14 (1.3)	0.4 (0.1, 1.7)	0.9 (0.2, 3.4)
History of cardiovascular disease, emphysema, or diabetes at enrollment				
Yes	150 (31.8)	363 (34.7)	0.9 (0.7, 1.1)	NA
No	322 (68.2)	684 (65.3)	Ref.	Ref.
Probable PTSD (checklist score ≥44) at enrollment				
Yes	124 (26.3)	174 (16.6)	1.8 (1.4, 2.4)	1.7 (1.3, 2.3)
No	345 (73.1)	860 (82.1)	Ref.	Ref.
Unknown	3 (0.6)	13 (1.2)	0.6 (0.2, 2.0)	1.0 (0.3, 3.9)

^a390 reports of non-melanoma or unspecified skin cancer were excluded from the analysis because non-melanoma or unspecified skin cancer are not reportable to state cancer registries.

^bPercentages may not total to 100 because of rounding.

^cAdjusting for variables that were listed in this table and were statistically significant or marginally significant in bivariate analyses

^dWave 2 refers to the first follow-up survey in 2006–2007.

^eUnknown age but indicated at ≥18 years at enrollment

OR = odds ratio

CI = confidence interval

Ref. = reference group

NA = not applicable

PTSD = posttraumatic stress disorder

diagnosis to data availability in the cancer registry and provide insight into the validity of cancers that are not reportable to the SCR. Studies examining false-negative reporting have cited factors such as being less informed about the diagnosis (possibly related to cultural differences in communication) or mistrust of health-care professionals or study interviewers.^{6,13} Researchers might improve reporting of health conditions by adopting other researchers' methods. Methods for enhancing response accuracy include providing a clear definition of the respondent's task, improving respondent's motivation, and facilitating cognitive processing.²² These methods are particularly important because African Americans had the highest overall cancer incidence and mortality rates between 2000 and 2004 among all races in the United States²³ and high incidence and mortality rates overall for other health conditions.²⁴⁻²⁶

Not completing a follow-up survey (Wave 2) was associated with false-negative reporting, which may have resulted from not having a second opportunity to report a cancer diagnosis or because of confounding factors associated with both nonparticipation and a cancer diagnosis. Those who did not complete the Wave 2 survey were more likely than those who completed Waves 1 and 2 to be older (≥ 65 years of age), have a lower annual household income ($< \$25,000$), or be current smokers, all of which are reported to be associated with false-negative reporting.^{2,6,8,11}

Limitations

This study had several limitations. First, one possible explanation for false-positive reporting is that self-reported cancer could not be verified in this study if the cancer was reported outside coverage of the 11 SCRs. Second, some cancers that were reported on the survey may not be reportable to SCRs. In our investigation of a subset of reports of cancer that were not confirmed by the SCRs, a large proportion of cancers confirmed by treating physicians were not reportable to an SCR. Thus, we would expect this discrepancy to result in some false-positive reports. Completeness and accuracy of data from the 11 SCRs used in this study have been reported to be relatively high,^{27,28} so this explanation for false-positive reports is unlikely. Third, self-report is subject to recall bias. Cancer was not covered by the September 11 Victim Compensation Fund until 2012,²⁹ five years after data used in this study were collected. As such, false-positive reports were unlikely to have been influenced by a desire to obtain financial support from the fund's existence.

CONCLUSION

Published WTC-related studies have largely relied on self-reported data. The findings in this study support the use of survey data for ongoing timely surveillance of adverse health conditions such as cancer. Self-reported cancer can be used to complement cancer linkage with cancer registries.

This study was supported by cooperative agreement #5U50/OH009739 and #1E11/OH009630 from the National Institute of Occupational Safety and Health (NIOSH), Centers for Disease Control and Prevention (CDC); #U50/ATU272750 from the Agency for Toxic Substances and Disease Registry, CDC, which included support from the National Center for Environmental Health, CDC, and the New York City Department of Health and Mental Hygiene (NYC DOHMH). The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of NIOSH or CDC. The authors assume full responsibility for analyses and interpretation of these data.

The authors acknowledge the support and collaborative effort of 11 state cancer registries for carrying out cancer record linkages, especially the following people: Allyn Ami, MPH, and Marilyn Kempster, MPH (California); Lou Gonsalves, PhD (Connecticut); Brad Wohler, MS (Florida); Pamela K. Agovino, MPH (New Jersey); Amy R. Kahn, MS (New York State); Chandrika Rao, PhD (North Carolina); Richard Knowlton, MS, MA, and BJ Mattson, MAE, MSTE (Ohio); Michelle Esterly, RHIA (Pennsylvania); Paul Betts, MS (Texas); and Mahesh Keitheri Cheteri, PhD (Washington). The authors also thank Lennon Turner, Renato Dasilva, and Rafaela Cruzado from the NYC DOHMH for their help in medical record confirmation; and Carolyn Greene, MD, Amy R. Kahn, and Sharon Perlman for their helpful comments.

This study was approved by the Institutional Review Board of NYC DOHMH. Cancer registry record linkages were approved by the respective institutional review boards of the state departments of health in California, Connecticut, Florida, Massachusetts, New York, North Carolina, Ohio, Pennsylvania, Texas, Washington, and Rutgers University (for New Jersey).

REFERENCES

1. Perlman SE, Friedman S, Galea S, Nair HP, Eros-Sarnyai M, Stellman SD, et al. Short-term and medium-term health effects of 9/11. *Lancet* 2011;378:925-34.
2. Schrijvers CT, Stronks K, van de Mheen DH, Coebergh JW, Mackenbach JP. Validation of cancer prevalence data from a postal survey by comparison with cancer registry records. *Am J Epidemiol* 1994;139:408-14.
3. Berthier F, Grosclaude P, Bocquet H, Faliu B, Cayla F, Machelard-Roumagnac M. Prevalence of cancer in the elderly: discrepancies between self-reported and registry data. *Br J Cancer* 1997;75:445-7.
4. Martin LM, Leff M, Calonge N, Garrett C, Nelson DE. Validation of self-reported chronic conditions and health services in a managed care population. *Am J Prev Med* 2000;18:215-8.
5. Yoshinaga A, Sasaki S, Tsugane S; JPHC Study Group (Japan Public Health Center). Sensitivity of self-reports of cancer in a population-based prospective study: JPHC Study Cohort I. *J Clin Epidemiol* 2001;54:741-6.
6. Desai MM, Bruce ML, Desai RA, Druss BG. Validity of self-reported cancer history: a comparison of health interview data and cancer registry records. *Am J Epidemiol* 2001;153:299-306.

7. Okura Y, Urban LH, Mahoney DW, Jacobsen SJ, Rodeheffer RJ. Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. *J Clin Epidemiol* 2004;57:1096-103.
8. Navarro C, Chirlaque MD, Tormo MJ, Perez-Flores D, Rodriguez-Barranco M, Sanchez-Villegas A, et al. Validity of self reported diagnoses of cancer in a major Spanish prospective cohort study. *J Epidemiol Community Health* 2006;60:593-9.
9. Kapp JM, Jackson-Thompson J, Petroski GF, Schootman M. Reliability of health-related quality-of-life indicators in cancer survivors from a population-based sample, 2005, BRFSS. *Public Health* 2009;123:321-5.
10. Colditz GA, Martin P, Stampfer MJ, Willett WC, Sampson L, Rosner B, et al. Validation of questionnaire information on risk factors and disease outcomes in a prospective cohort study of women. *Am J Epidemiol* 1986;123:894-900.
11. Parikh-Patel A, Allen M, Wright WE; California Teachers Study Steering Committee. Validation of self-reported cancers in the California Teachers Study. *Am J Epidemiol* 2003;157:539-45.
12. Freedman DM, Sigurdson AJ, Doody MM, Love-Schnur S, Linet MS. Comparison between cancers identified by state cancer registry, self-report, and death certificate in a prospective cohort study of US radiologic technologists. *Int J Epidemiol* 2006;35:495-7.
13. Manjer J, Merlo J, Berglund G. Validity of self-reported information on cancer: determinants of under- and over-reporting. *Eur J Epidemiol* 2004;19:239-47.
14. Farfel M, DiGrande L, Brackbill R, Prann A, Cone J, Friedman S, et al. An overview of 9/11 experiences and respiratory and mental health conditions among World Trade Center Health Registry enrollees. *J Urban Health* 2008;85:880-909.
15. Murphy J, Brackbill RM, Thalji L, Dolan M, Pulliam P, Walker DJ. Measuring and maximizing coverage in the World Trade Center Health Registry. *Stat Med* 2007;26:1688-701.
16. Brackbill RM, Hadler JL, DiGrande L, Ekenga CC, Farfel MR, Friedman S, et al. Asthma and posttraumatic stress symptoms 5 to 6 years following exposure to the World Trade Center terrorist attack. *JAMA* 2009;302:502-16.
17. Centers for Disease Control and Prevention (US). National Program of Cancer Registries (NPCR): Registry Plus™ Link Plus [cited 2016 Jan 20]. Available from: <http://www.cdc.gov/cancer/npcr/tools/registryplus/lp.htm>
18. National Cancer Institute (US), Surveillance, Epidemiology, and End Results Program. SEER site recode B ICD-O-3 (12/19/2003) definition [cited 2016 Jan 20]. Available from: http://seer.cancer.gov/siterecode_b/icdo3_d12192003
19. SAS Institute, Inc. SAS®: Version 9.2. Cary (NC): SAS Institute, Inc.; 2010.
20. Bergmann MM, Calle EE, Mervis CA, Miracle-McMahill HL, Thun MJ, Heath CW. Validity of self-reported cancers in a prospective cohort study in comparison with data from state cancer registries. *Am J Epidemiol* 1998;147:556-62.
21. Cho LY, Kim CS, Li L, Yang JJ, Park B, Shin A, et al. Validation of self-reported cancer incidence at follow-up in a prospective cohort study. *Ann Epidemiol* 2009;19:644-6.
22. Del Boca FK, Noll JA. Truth or consequences: the validity of self-report data in health services research on addictions. *Addiction* 2000;95 Suppl 3:S347-60.
23. National Cancer Institute. Cancer health disparities. 2008 [cited 2015 Nov 12]. Available from: <http://www.cancer.gov/about-nci/organization/crhd/cancer-health-disparities-fact-sheet>
24. Williams DR. Race, socioeconomic status, and health: the added effects of racism and discrimination. *Ann N Y Acad Sci* 1999;896:173-88.
25. Adams PF, Kirzinger WK, Martinez ME. Summary health statistics for the U.S. population: National Health Interview Survey, 2011. *Vital Health Stat* 10 2012(255).
26. Centers for Disease Control and Prevention (US). Asthma in the US [cited 2015 Nov 12]. Available from: <http://www.cdc.gov/vitalsigns/asthma>
27. Thoburn KK, German RR, Lewis M, Nichols PJ, Ahmed F, Jackson-Thompson J. Case completeness and data accuracy in the Centers for Disease Control and Prevention's National Program of Cancer Registries. *Cancer* 2007;109:1607-16.
28. North American Association of Central Cancer Registries. U.S. registries certified in 2011 for 2008 incidence data [cited 2016 Jan 20]. Available from: <https://www.naaccr.org/Certification/USCert2008.aspx>
29. Department of Health and Human Services (US), Centers for Disease Control and Prevention. World Trade Center Health Program; addition of certain types of cancer to the list of WTC-related health conditions. *Fed Reg* 2012;77:56138-68.